THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 06_11

Corrado Benassi
Dipartimento di Scienze Economiche
Alma Mater Studiorum - Università di Bologna
and
The Rimini Centre for Economic Analysis

A NOTE ON CONVEX TRANSFORMATIONS AND THE FIRST ORDER APPROACH

This version:
January 2011

A Note on Convex Transformations and the First Order Approach

Corrado Benassi
Dipartimento di Scienze Economiche
Alma Mater Studiorum - Università di Bologna
and
The Rimini Center for Economic Analysis

Abstract The first order approach to solving the standard one-dimensional principal-agent model is conditional upon the relevant stochastic production function obeying two noteworthy restrictions: that the Likelihood Ratio be monotonically increasing in output, and that the distribution function be convex in effort. It is usually claimed that such conditions are very restrictive, as very few of the standard probability distributions satisfy both properties. The purpose of this note is to show that this lack of generality should not be seen as a problem, since some simple convexifying transformations are available that enable one to work with proper distributions with the required properties. *JEL Classification no*: D86.

Keywords Principal agent problem, first order approach.

Corresponding author C.Benassi, Dipartimento di Scienze Economiche, Università di Bologna, Piazza Scaravilli 2, 40126 Bologna, Italy. Phone: +39 051 2098667; Fax: +39 051 221968; e-mail: corrado.benassi@unibo.it

1 Introduction

It is well known that the first order approach to solving the standard one-dimensional principal-agent model is conditional upon the relevant stochastic production function obeying two noteworthy restrictions: the Monotone Likelihood Ratio Property and that the distribution function be convex in effort. It is usually claimed that such conditions are very restrictive, as very few of the standard probability distributions – like the Normal, Beta, Chi square, F, Weibull, etc. – satisfy both properties (e.g., Jewitt, 1988; Li Calzi and Spaeter, 2003).

The purpose of this note is to show that this lack of generality should not be seen as a problem, since some simple convexifying transformations are available, that enable one to work with suitably transformed standard distributions, in such a way that some key properties (e.g., unimodality or symmetry) are preserved consistently with those required for contract-theoretic modeling.

2 The basic model

The standard formulation of the continuous principal-agent model with separable utility[1] takes the general form

$$\max_{w(\cdot),a^*} \int_{\mathcal{X}} v\left[x - w(x)\right] f(x,a^*)dx \tag{1.a}$$

subject to

$$\int_{\mathcal{X}} u\left(w\left(x\right)\right) f(x,a^*) - c\left(a^*\right) \geq \overline{u} \tag{1.b}$$

$$a^* \in \arg\max_{a} \int_{\mathcal{X}} u\left(w\left(x\right)\right) f(x,a)dx - c\left(a\right) \tag{1.c}$$

where x denotes output and a effort; $v\left(\cdot\right)$ is the principal's payoff, and $u\left(\cdot\right) - c(a)$ the agent's, with $c(a)$ increasing convex and $w\left(\cdot\right)$ the payment schedule from the principal to the agent; $\overline{u}$ denotes the agent's reservation utility, and it is usually assumed that both u and v are increasing concave functions.

[1]Non separable utility is much harder to work with. An extension of the FOA to the nonseparable utility case is provided by Alvi (1997).

Letting subscripts denote derivatives, $f(x,a) = F_x(x,a)$ denotes the strictly positive density of

$$F : \mathcal{X} \times \mathcal{A} \rightarrow [0,1]$$

which gives the parametrized distribution $F(x,a)$ of observable (and verifiable) output $x \in \mathcal{X}$, given the agent's hidden action $a \in \mathcal{A} \subset \mathbb{R}_+$. It is standardly assumed that the support $\mathcal{X}$ is a compact interval independent of a; that F is continuously differentiable at least twice; and that the agent's effort exerts a positive effect on output in the sense of first order stochastic dominance, that is

$$F_a(x,a) \leq 0 \tag{2}$$

for all $a \in \mathcal{A}$.[2]

Problem (1) embodies the idea of the principal maximizing her utility subject to the agent's participation constraint (1.b) – whereby he gains at least his reservation utility; and the agent's incentive compatibility constraint (1.c) – whereby his (unobservable) chosen action is consistent with the principal's objective.

Problem (1) is in general very difficult to solve, so that following Mirrlees (1975) the so-called First Order Approach (FOA) is usually invoked:[3] that is, constraint (1.c) is substituted by

$$\int_{\mathcal{X}} u\left(w\left(x\right)\right) f_a(x,a^*)dx - c_a\left(a^*\right) = 0 \tag{3}$$

i.e., by the requirement that a^* be the a stationary point of the agent's expected utility.[4] Subsequent literature (Holmström,1979; Rogerson, 1985) established that the FOA, though not valid in general, it is so if the following twin conditions are satisfied

$$F_{aa}(x,a) \geq 0 \text{ for all } x \in \mathcal{X} \tag{4.a}$$
$$\alpha_x\left(x,a\right) \geq 0 \text{ for all } x \in \mathcal{X} \tag{4.b}$$

[2] The set $\mathcal{A}$ is sometimes assumed to be an open interval, so as to characterize optima by interior maxima (e.g., Jewitt, 1988, p.1179)

[3] An important complementary approach is provided by Grossman and Hart (1983).

[4] A generalized Lagrangean approach to the FOA that includes multipliers for second order conditions is provided by Araujo and Moreira (2001).

Condition (4.a) is usually known as the CDF property (effort-convexity of the distribution function), while (4.b) is known as the MRLP (Monotone Likelihood Ratio Property), where

$$\alpha(x,a) = \frac{f_a(x,a)}{f(x,a)}$$

is the likelihood ratio, assumed to be monotonically increasing in x.

While the MRLP is generally looked at as a non-controversial assumption, the CDF property is usually considered very much restrictive. Indeed, though one reasonable implication of the CDF property is that of decreasing marginal (expected) productivity of effort,[5] very few distributions seem to share this property, a property which moreover does have one disturbing feature highlighted by Jewitt (1988, p.1177):

[consider the case where o]utput is subject to a simple additive disturbance ε with distribution function F, and effort, a, is measured in output terms. So, realized output is given by $x = a + \varepsilon$, and this has distribution function $F(x-a)$ which is only convex in effort if ε has an increasing density! Hence, an apparently natural case does not fit the condition.

3 Two simple convexifying transformations

The question we ask is whether there exist convexifying mappings which may transform some given distribution into one satisfying the properties required by the FOA, *viz* the MRLP and the CDF property, while retaining some features (like unimodality and symmetry) which make sense in a variety of environments.[6]

Generally speaking, the issue of generating distributions by appropriate transformations is addressed by the statistical literature within the framework of distribution *systems* (e.g., Johnson *et al.*,1994, ch.12), which may

[5] Since expected output is $\mu(a) = 1 - \int_{\mathcal{X}} F(x,a)dx$, $\mu_{aa}(a) < 0$ follows from $F_{aa}(x,a) \geq 0$ for all x.

[6] In this sense our approach is somehow complementary to that of Li Calzi and Spaeter (2003), who develop general formulations for classes of densities satisfying the MRL and the CDF properties, such that " these classes are generic and encompass a large number of specific functional forms." (p.168).

also be relevant to the problem at hand. For example, the so called S-system (Voit, 1992) collects distributions obeying the differential equation

$$\frac{dF}{dx} = \left[F^b - F^c\right]\gamma \tag{5}$$

with $\gamma > 0$ and $b < c$. This family includes the logistic distribution ($b = 1$, $c = 2$), which can be shown to obey both the MLR and the CDF properties for an appropriate parameter choice.[7]

In this note, however, we shall focus on two more specific convex transformations which in principle are meant to be applicable to any distribution.

3.1 Case 1

This case relies on the idea of appropriately parametrizing some "core" output distribution $G(x) : \mathcal{X} \to [0,1]$. If one lets $\varphi : [0,1] \times \mathcal{A} \to [0,1]$, any given a induces a distribution F such that $F : \mathcal{X} \times \mathcal{A} \to [0,1]$:

$$F(x,a) = \varphi\left(G(x), a\right)$$

It is then just a matter of simple differentiation to prove the following:

Proposition 1 *Let φ have the following properties for all $x \in \mathcal{X}$ and $a \in \mathcal{A}$: (i) $\varphi(0,a) = 0$, $\varphi(1,a) = 1$; (ii) $\varphi_1(\cdot, a) \geq 0$; (iii) $\frac{\partial}{\partial x}\frac{\varphi_{12}}{\varphi_1} \geq 0$; (iv) $\varphi_{22} \geq 0$. Then F is a distribution obeying both the MRLP and the CDF property.*

We present two examples:

Example 1 One example which arises quite naturally is the power function. Take *any* distribution $G : \mathcal{X} \to [0,1]$, in principle satisfying neither CDF nor MRLP: then if one defines $F(x,a) = G(x)^a$, F is a proper distribution for any $a > 0$, which does satisfy both the CDF property and the MRLP. Indeed, it

[7] The (unimodal) logistic density is $\ell(x,a;s) = (4s)^{-1}\operatorname{sech}^2\left(\frac{x-a}{2s}\right)$ where a is the mean and s a dispersion parameter (Johnson et al., 1995, p.115). By normalizing the support as $\mathcal{X}$ =[0,1], one obtains one $f(x,a;s)$: e.g., for $s = 2$ this has the required properties for any $a > 1$.

is readily seen that $F_{aa}(x,a) = G(x)^a \ln^2 G(x) > 0$ for all x, while the density is $f(x,a) = aG(x)^{a-1} g(x)$ and the likelihood ratio is

$$\alpha(x,a) = \frac{f_a(x,a)}{f(x,a)} = \tfrac{1}{a} + \ln G(x) \tag{6}$$

which is obviously increasing in x. As an instance, when the "core" distribution G is the standardized normal $N(0,1)$, one has

$$f(x,a) = (2\pi)^{-\frac{1}{2}a}\, e^{-\frac{1}{2}ax^2}$$

which is symmetric and unimodal for any positive a.

Example 2 Another example is given by the following transformation. Take again any $G : \mathcal{X} \to [0,1]$ and define

$$F(x,a) = \frac{(ae)^{G(x)} - 1}{ae - 1} \tag{7}$$

for $\mathcal{A} = (1,\infty)$. One can check that $F_{aa}(x,a)$ is positive for all $x \in \mathcal{X}$ and $a > 1$; the likelihood ratio is

$$\alpha(x,a) = \frac{f_a(x,a)}{f(x,a)} = \frac{1}{a}\left(G(x) - \frac{1 + ea\ln a}{(ae-1)(1+\ln a)}\right)$$

which is clearly monotonically increasing in x.[8]

3.2 Case 2

A somewhat different approach relies on applying a suitable transformation to the distribution's support. Let the distribution $H : \mathcal{Y} \to [0,1]$, and let $\psi : \mathcal{X} \times \mathcal{A} \to \mathcal{Y}$ so that any given a induces a distribution $F : \mathcal{X} \times \mathcal{A} \to [0,1]$ such that

$$F(x,a) = H(\psi(x,a))$$

Then the following can be easily established:

[8] In this connection one should perhaps mention a third example, where $F(x,a) = [\exp\{G(x)^a\} - 1] / (e-1)$, which gives the same F as definition (7) for $a = 1$, and can be seen as a transformation Example 1. One can easily check that also in this case $F_{aa}(x,a)$ is positive and the likelihood ratio is increasing.

Proposition 2 *Let $\psi(x,a)$ and $h(\psi) = H_\psi(\psi)$ have the following properties for all $x \in \mathcal{X}$ and $a \in \mathcal{A}$: (i) $\psi_1 > 0$; (ii) $\psi_2 < 0$; (iii) $\frac{\partial}{\partial x}\left(\frac{h_\psi(\psi)}{h(\psi)}\psi_a + \frac{\psi_{ax}}{\psi_x}\right) > 0$; (iv) $\frac{h_\psi(\psi)}{h(\psi)}\psi_a^2 + \psi_{aa} > 0$. Then F is a distribution obeying both the MRLP and the CDF property.*

Again we consider here two examples, the presentation of which is simplified by characterizing densities via their Esteban elasticity:[9] for any generic density $p(\cdot)$ the latter is defined as

$$\pi^p(x) = \lim_{h\to 0}\frac{d\log\left(\frac{1}{\mu}\int_x^{x+h} zp(z)dz\right)}{d\log x} = 1 + \frac{xp_x(x)}{p(x)} \tag{8}$$

where μ is the distribution mean.

Example 3 A natural case in point is the well known Jewitt formulation, where $\psi(x,a) = x - a$: which however – as Jewitt (1988) emphasizes – is consistent with both the required properties only at the cost of f being an increasing density. Arguably, however, the additive structure is somehow conditional on the chosen units.[10] Suppose then that $\psi(x,a) = e^{x-a}$:[11] it is readily seen that $\alpha(x,a) = -\pi^h(\psi)$ (so that the MLRP is satisfied if $\pi^h(\cdot)$ is monotonically decreasing), and that the CDF property requires $\pi^h(\psi) > 0$. Hence, both properties are satisfied by all H such that π^h is positive and decreasing, which holds for a variety distributions, and is consistent with unimodality for an appropriate choice of parameters. For example, let H be the Gamma distribution normalized over, say, $[1,4]$:

$$H(\psi;\beta,\gamma) = \frac{\Gamma(\beta,1/\gamma) - \Gamma(\beta,\psi/\gamma)}{\Gamma(\beta,1/\gamma) - \Gamma(\beta,4/\gamma)} \tag{9}$$

[9] Esteban (1986) defines this in the context of income distribution as "income share elasticity": π^f is shown to be a legitimate representation of f in that it stands one-to-one with f itself. It should be noticed that "the Pareto, Gamma and Normal density functions correspond to constant, linear and quadratic elasticities, respectively" (p.442). Note also that, if the density takes the form $f(x,a)$, with a some parameter, the condition $\pi_a^f(x,a) > 0$ amounts to the MLRP (Benassi and Chirco, 2006).

[10] As Holmström and Hart (1987, p.81) argue when discussing the possibilty of linear sharing rules, "the connection between x as physical output and as statistical information is very tenuous [...]; all that matters is the distribution of the posterior (or likelihood ratio) as a function of the agent's action".

[11] This amounts to the log version of the Jewitt case, as $\ln\psi = x - a$.

where $\Gamma(\cdot,\cdot)$ is the incomplete Gamma function, with $\beta > 1$ and $\gamma > 0$.[12] One can check that

$$\pi^h(\psi;\beta,\gamma) = \beta - \frac{\psi}{\gamma} \tag{10}$$

monotonically decreasing and positive for $\psi \leq \beta\gamma$: letting $\psi(x,a) = e^{x-a}$ one has

$$f(x,a;\beta,\gamma) = \frac{\gamma^{-\beta}\exp\left\{(\beta-1)(x-a) - \gamma^{-1}e^{x-a}\right\}}{\Gamma\left(\beta,\frac{1}{\gamma}\right) - \Gamma\left(\beta,\frac{4}{\gamma}\right)}$$

and likelihood ratio

$$\alpha(x,a;\beta,\gamma) = 1 - \beta + \frac{1}{\gamma}e^{x-a} \tag{11}$$

which is clearly monotonically increasing in x. Suppose for instance that $(\beta,\gamma) = (2,3)$:[13] then unimodality is consistent with this suitably transformed additive structure.

Example 4 In a somewhat similar vein, one can have $\psi(x,a) = x/a$,[14] in which case $\alpha(x,a) = -\pi^h(\psi)/a$ and

$$sign\{F_{aa}(x,a)\} = sign\left\{\left(\pi^h(\psi)+1\right)\frac{x}{a^3}\right\} \tag{12}$$

so that, again, for $a > 0$ any distribution H with positive and decreasing Esteban elasticity will do.

4 Concluding remarks

The twin conditions of Monotone Likelihood Ratio Property and Convexity of the Distribution are often claimed to be too stringent for the First Order

[12] That is, $\Gamma(x,y) = \int_x^\infty e^{-t}t^{y-1}dt$, such that $\Gamma(0,y) = \Gamma(y)$.

[13] Clearly, we are imposing $\psi = e^{x-a} \in [1,4]$. The mode is $x^m = a + \ln[(\beta-1)\gamma]$, such that $e^{x^m-a} = (\beta-1)\gamma$; since π^h positive and decreasing requires $\beta\gamma > 4$, the parameter constraint is $\gamma \in \left(\frac{4}{\beta},\frac{4}{\beta-1}\right)$.

[14] This clearly amounts to interpreting Jewitt's formulation in log terms.

Approach to the solution of the standard principal-agent problem to be viable in many applications, as most distributions currently in use do not satisfy them. In this short note we have shown that this may in fact not be so, as some convexifying transformations are available which allow one to work with suitably transformed standard distributions.

References

[1] Alvi, E. (1997): First-Order Approach to Principal-Agent Problems: A Generalization, *The Geneva Papers on Risk and Insurance Theory*, **22**, 59-65.

[2] Araujo, A. and H.Moreira (2001): A general Lagrangian approach for non-concave moral hazard problems, *Journal of Mathematical Economics*, **35**, 17-39.

[3] Benassi, C. and A.Chirco (2006): Income share elasticity and stochastic dominance, *Social Choice and Welfare*, **26**, 511-25.

[4] Esteban, J. (1986): Income Share Elasticity and the Size Distribution of Income, *International Economic Review*, **27**, 439-44.

[5] Grossman, S. and O.Hart (1983): An Analysis of the Principal-Agent Problem, *Econometrica*, **51**, 7-45.

[6] Hart, O. and B.Holström (1987): The Theory of Contracts, in T.F.Bewley (ed.), *Advances in Economic Theory. Fifth World Congress.* Cambridge University Press, Cambridge and New York, 71-156.

[7] Holström, B. (1979): Moral Hazard and Observability, *The Bell Journal of Economics*, **10**, 74-91.

[8] Jewitt, I. (1988): Justifying the first-order approach to principal-agent problems, *Econometrica*, **56**, 1117-1190.

[9] Johnson, N.L., S.Kotz and N.Balakrishnan (1994): *Continuous Univariate Distributions*, vol.1. J.Wiley & Sons, New York and Toronto.

[10] Johnson, N.L., S.Kotz and N.Balakrishnan (1995): *Continuous Univariate Distributions*, vol.2. J.Wiley & Sons, New York and Toronto.

[11] Li Calzi, M. and S.Spaeter (2003): Distributions for the first-order approach to principal-agent problems, *Economic Theory*, **21**, 167-173.

[12] Mirrlees, J. (1975): The Theory of Moral Hazard and Unobservable Behaviour: Part 1, *The Review of Economic Studies* (1999), **66**, 3-21.

[13] Rogerson, W.P. (1985): The First-Order Approach to Principal-Agent Problems, *Econometrica*, **53**, 1357-1367.

[14] Voit, E.O. (1992): The *S*-Distribution: A tool for approximation and classification of univariate, unimodal probability distributions, *Biometrical Journal*, **34**, 855-78.